dr. Mantas Lukoševičius
mantas.lukosevicius@ktu.lt
https://mantas.info/
2022.09.13

ktu faculty of informatics 1922

Advanced Machine Learning
P176M010

2. Introduction to Machine Learning

Outline

- What is Machine Learning (ML)
  - Definition, history, ML and AI
  - ML and data science
- ML types and basic setup
  - Optimal decision boundary
  - Curse of dimensionality

The direct goals of this course

- Be able to apply ML methods to solve real world problems
- To master a “toolbox” of effective ML methods
- Be able to select parameters, train, evaluate ML methods
- To know the classes of ML methods, their principles, pros & cons
- Be able to understand the data, prepare them for ML
- To understand the fundamental ML principles, notions, “laws“

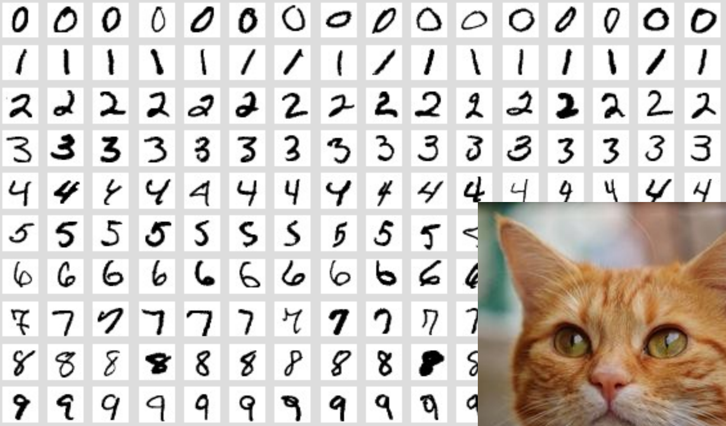
Machine Learning: Use Case

- When we know and have examples of **what** the solution looks like but **not how** to solve the problem – no exact algorithm
- Or we want to make sense of the (huge) data

+ Expands applicability of computers to “soft” domains

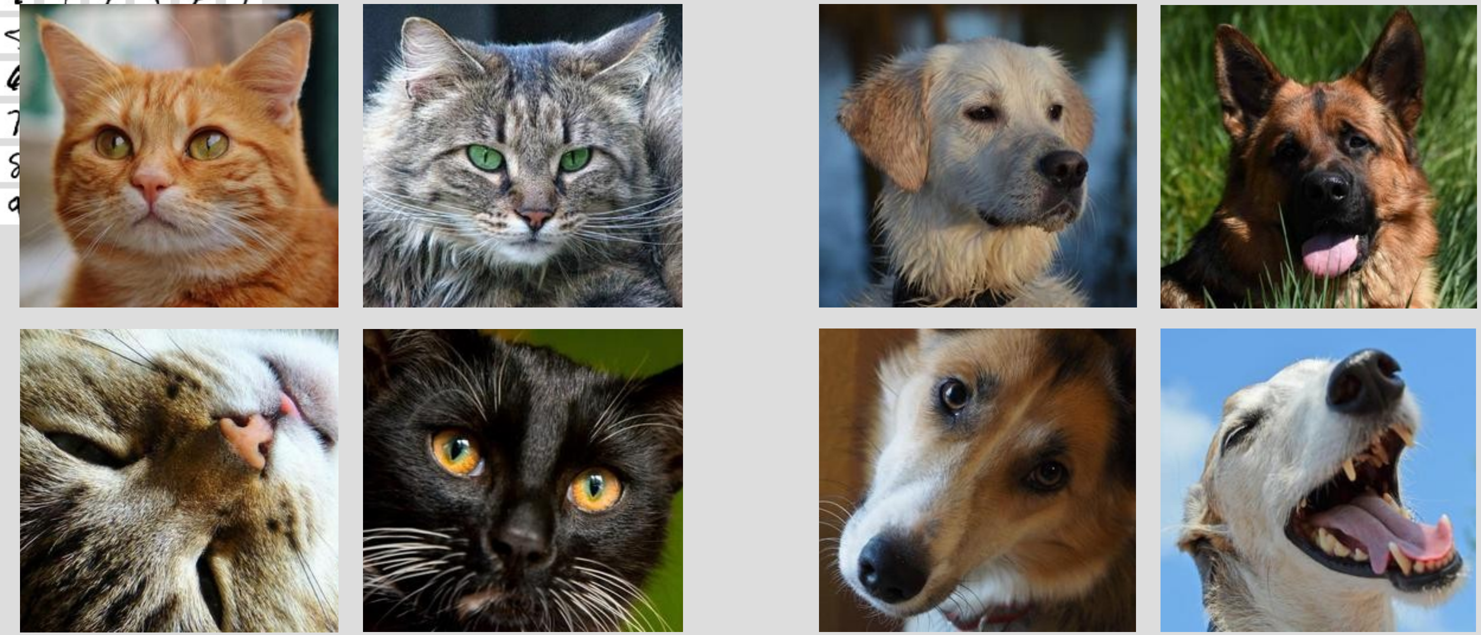
– We need to have representative data

Sample ML problems



https://en.wikipedia.org/wiki/MNIST_database

Recognize cats and dogs in pictures

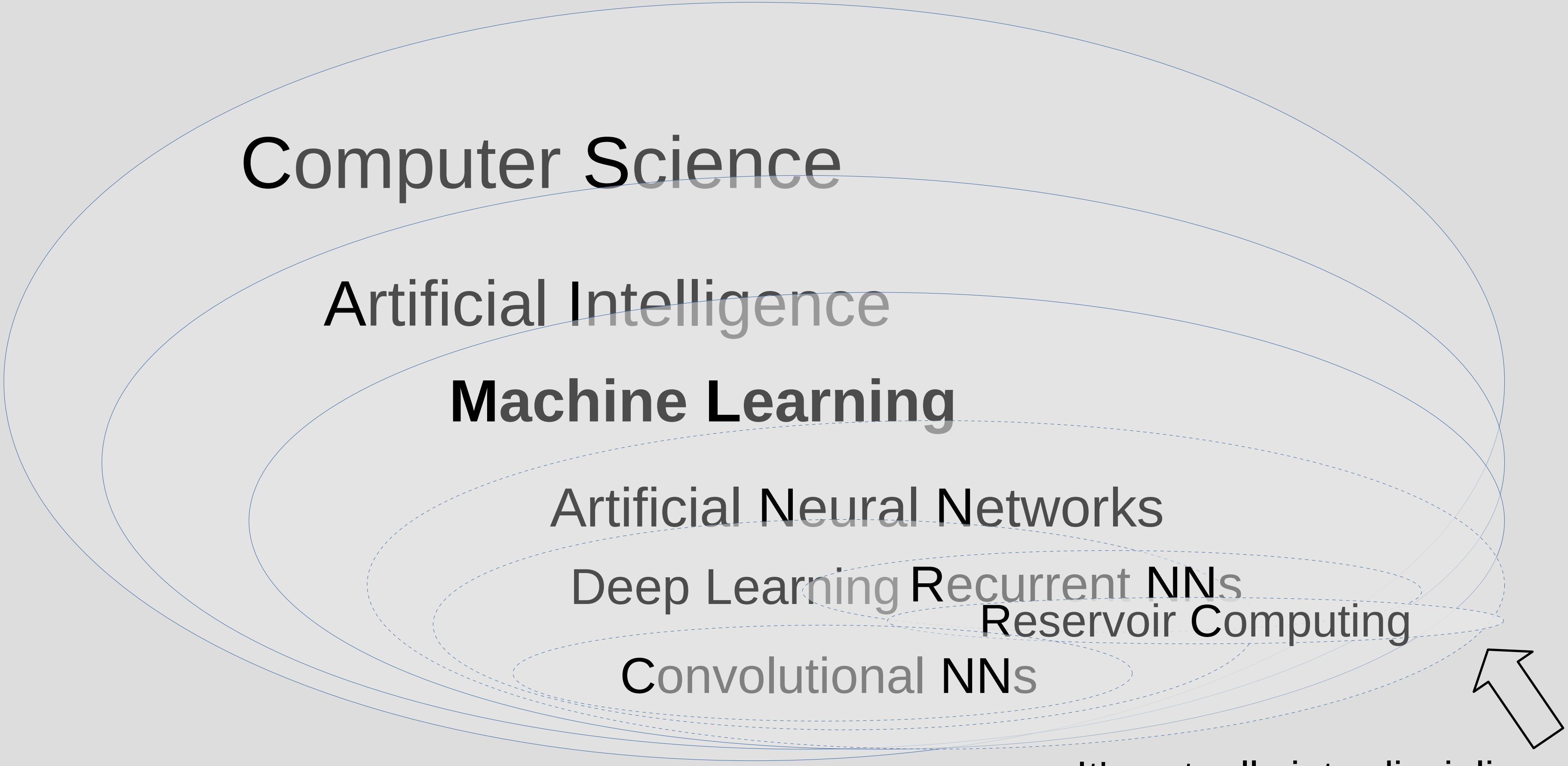


Machine Learning – Definition

- Arthur Samuel (1959):
  “*Machine Learning: Field of study that gives computers the ability to learn without being explicitly programmed.*”

  – programs that improve with experience

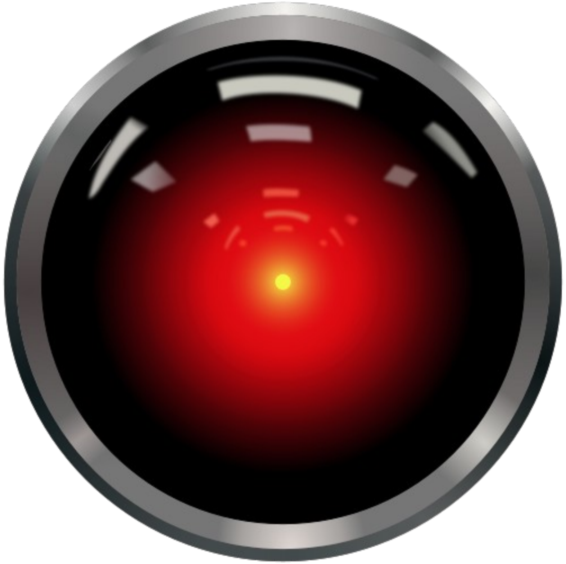
The Field



It's actually interdisciplinary!

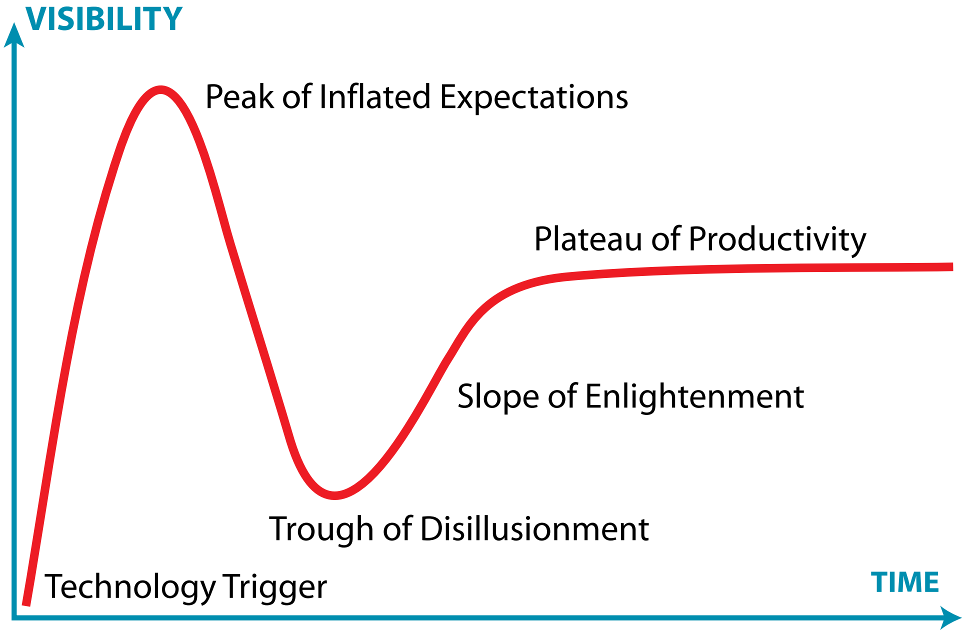
Artificial Intelligence: bits of History

- Math | very long ago
- Philosophical logics | 1st millennium BC
- Mechanical calculators | 17th century
- Mathematical logics | 20th century
- Computers | 1946
- Birth of AI field, golden years | 1956
- “AI winter” | 1974 ~ 2010?
- Narrow solutions behind the scenes | 1993 – …
- "Deep Learning" | 2000 – …

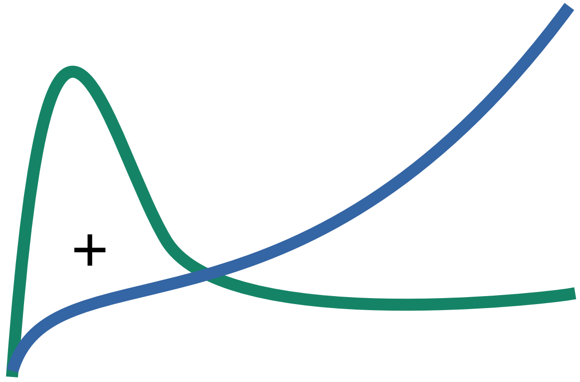


Pic. by Cryteria

Technology Hype Cycle



$\stackrel{?}{=}$



Symbolic (“Good Old-Fashioned”) AI vs. Machine (Deep) Learning?

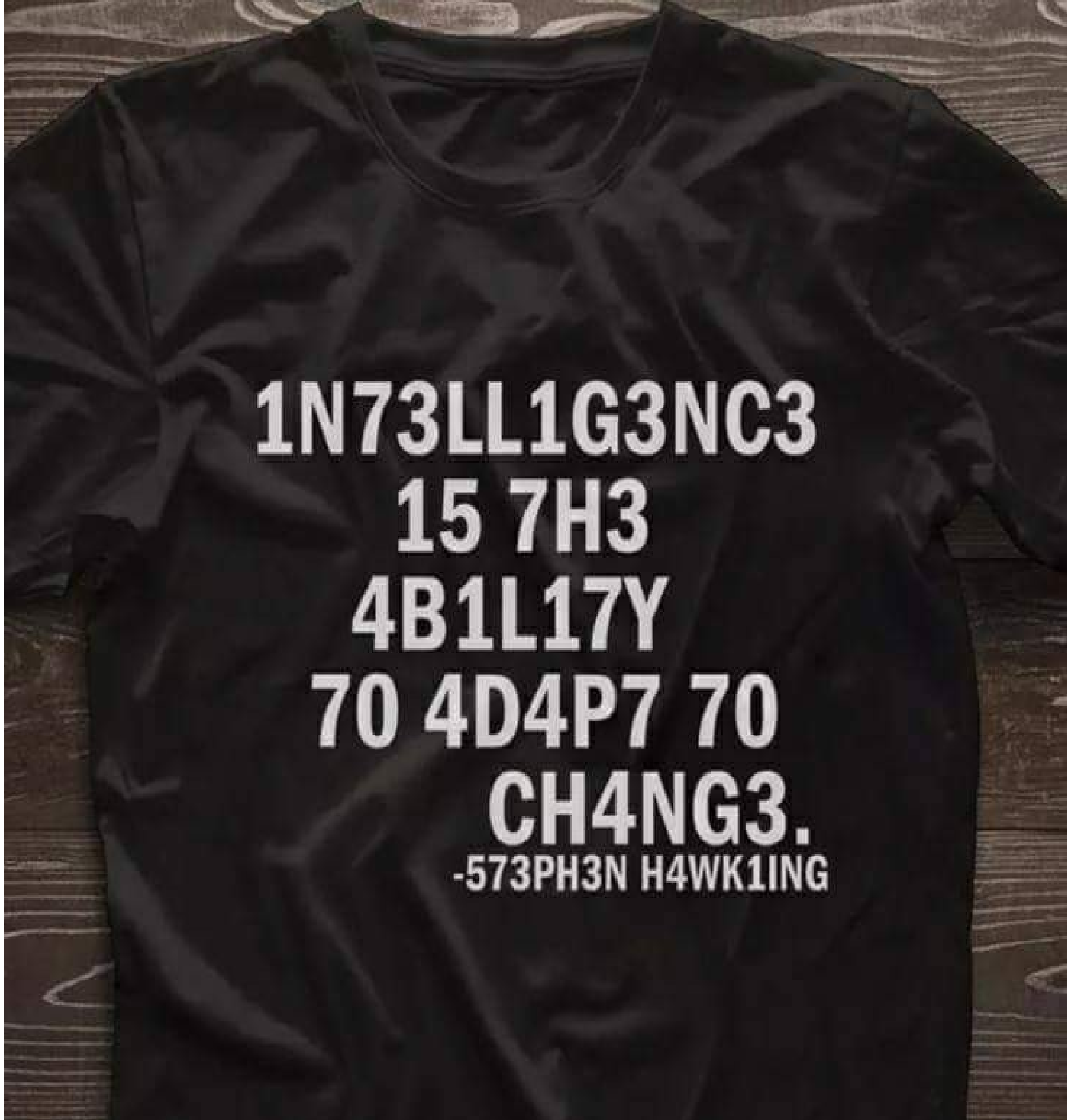
https://en.wikipedia.org/wiki/Hype_cycle

by Gartner, Inc

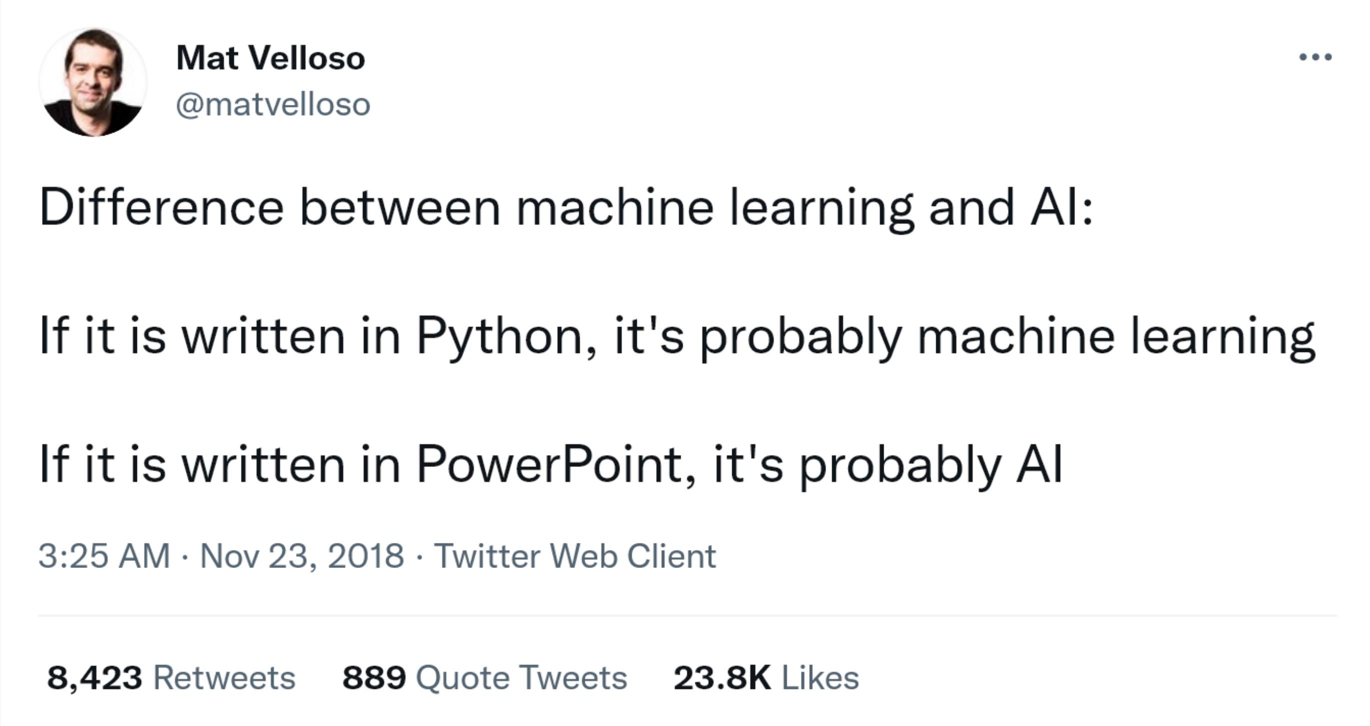
https://en.wikipedia.org/wiki/Symbolic_artificial_intelligence

Machine Learning and AI

- Was there from the beginning
  - Arthur Samuel: Learning checkers player – 1952
  - Frank Rosenblatt: Perceptron – 1957, ...
- Relative importance only increases
- Machine Learning captures important aspects of intelligence
  - **Adaptivity** to the environment
  - How intelligence **comes into being**
- Influenced by other fields
- Machine learning is becoming increasingly mathematically rigorous
  - = Statistics + computer science?
  - … and/or deep ;)


1N73LL1G3NC3
15 7H3
4B1L17Y
70 4D4P7 70
CH4NG3.
-573PH3N H4WK1ING

ML (Deep Learning) dominates modern AI



Mat Velloso
@matvelloso

Difference between machine learning and AI:

If it is written in Python, it's probably machine learning

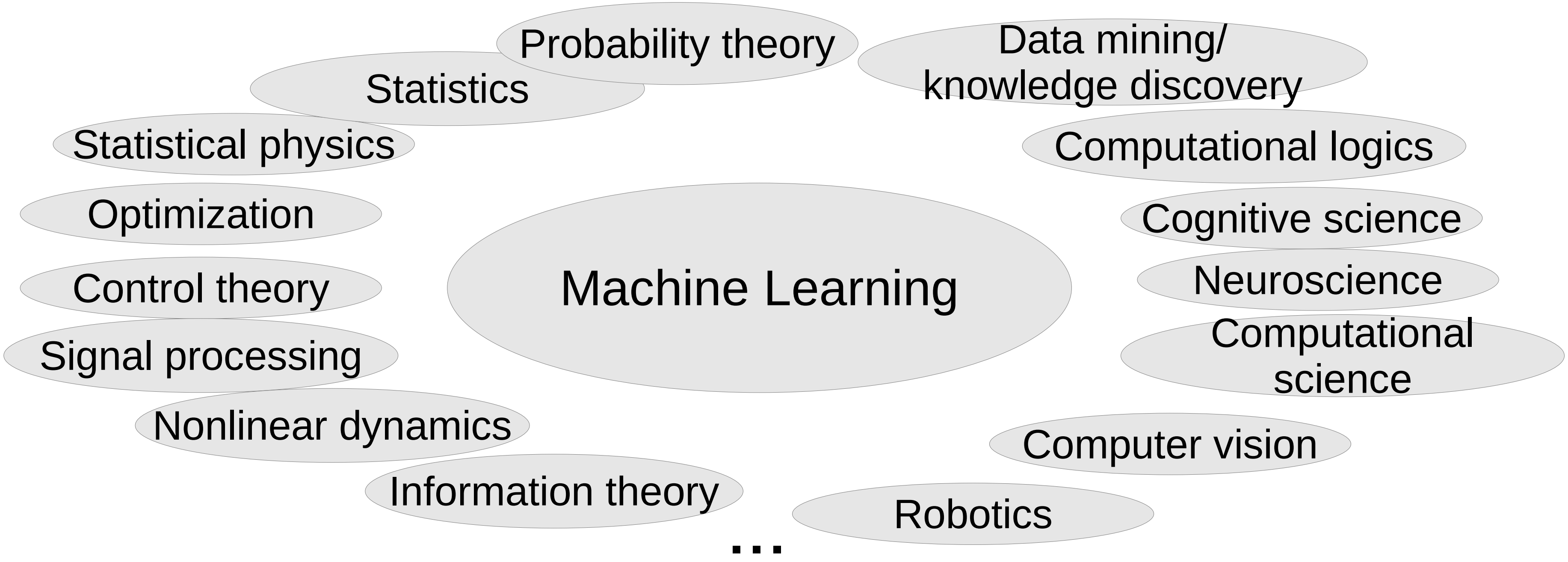
If it is written in PowerPoint, it's probably AI

3:25 AM · Nov 23, 2018 · Twitter Web Client

8,423 Retweets **889** Quote Tweets **23.8K** Likes

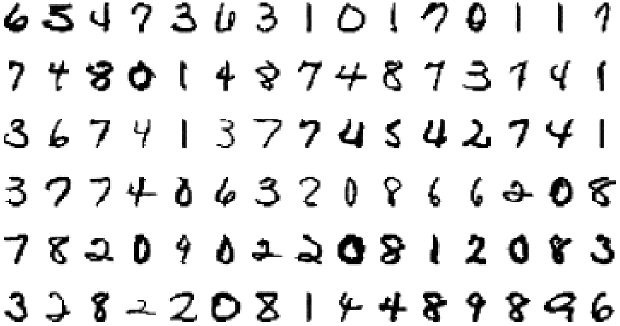
https://twitter.com/matvelloso/status/1065778379612282885

Overlapping/related fields



Classical Task: Digit recognition

- USPS



Pic from: Z. Zhenga, J Yang, 06

- MNIST

https://en.wikipedia.org/wiki/MNIST_database

- Google house numbers

http://ufldl.stanford.edu/housenumbers/

Machine Learning Applications

- Web search ranking, feed composition
- Automated recommendations, ads
- Fraud detection, cybersecurity, spam filters
- Lower level control: auto-focus, engines, data transmission, etc.
- Face, movement recognition, biometrics
- Object recognition
- Self-driving cars
- Speech, query intent recognition
- Text sumarization, generation
- Music, art generation, style transfer
- Medical applications: diagnosis, monitoring
- Bioinformatics
- Brain-computer interfaces
- Robotics
- Games
- Revolutionizes many other fields...

Text generation demo

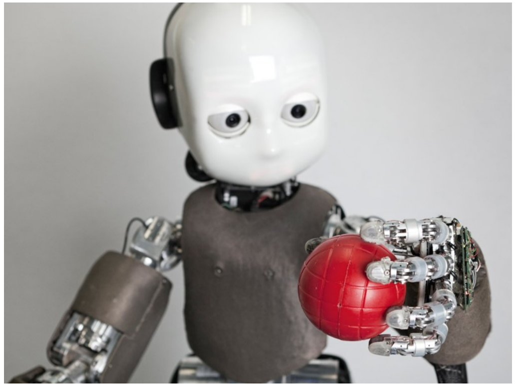
- Text generator powered by GPT-J-6B model at https://6b.eleuther.ai/
  (GPT-3 is too expensive for such a demo)

A new Advanced Machine Learning course is taught to the Master's students in informatics at Kaunas University of Technology. The new course should allow you to: 1) to increase your knowledge of machine learning, 2) to master advanced machine learning techniques and 3) to have practical experience in solving problems using advanced ML techniques. It consists of theoretical lectures, hands-on projects, and optional practice of practical problems. We will teach advanced techniques on the following topics: neural networks, decision trees, support vector machines, principal component analysis, logistic regression, neural-network-based sequence prediction, recurrent neural networks, reinforcement learning, reinforcement learning, reinforcement learning with experience replay, deep reinforcement learning and genetic algorithms.

This course will have practical exercises, using the datasets obtained from...

Limitations of Machine Learning Compared to Natural

- We have tons of context knowledge about the world, computer only sees a piece of data
- Algorithms need lots of labeled data „to get the idea“
- Computational power might still be too weak
- Understanding of the brain is limited
- Embeddedness and embodiment
- Active learning
- ...



https://icub.iit.it/

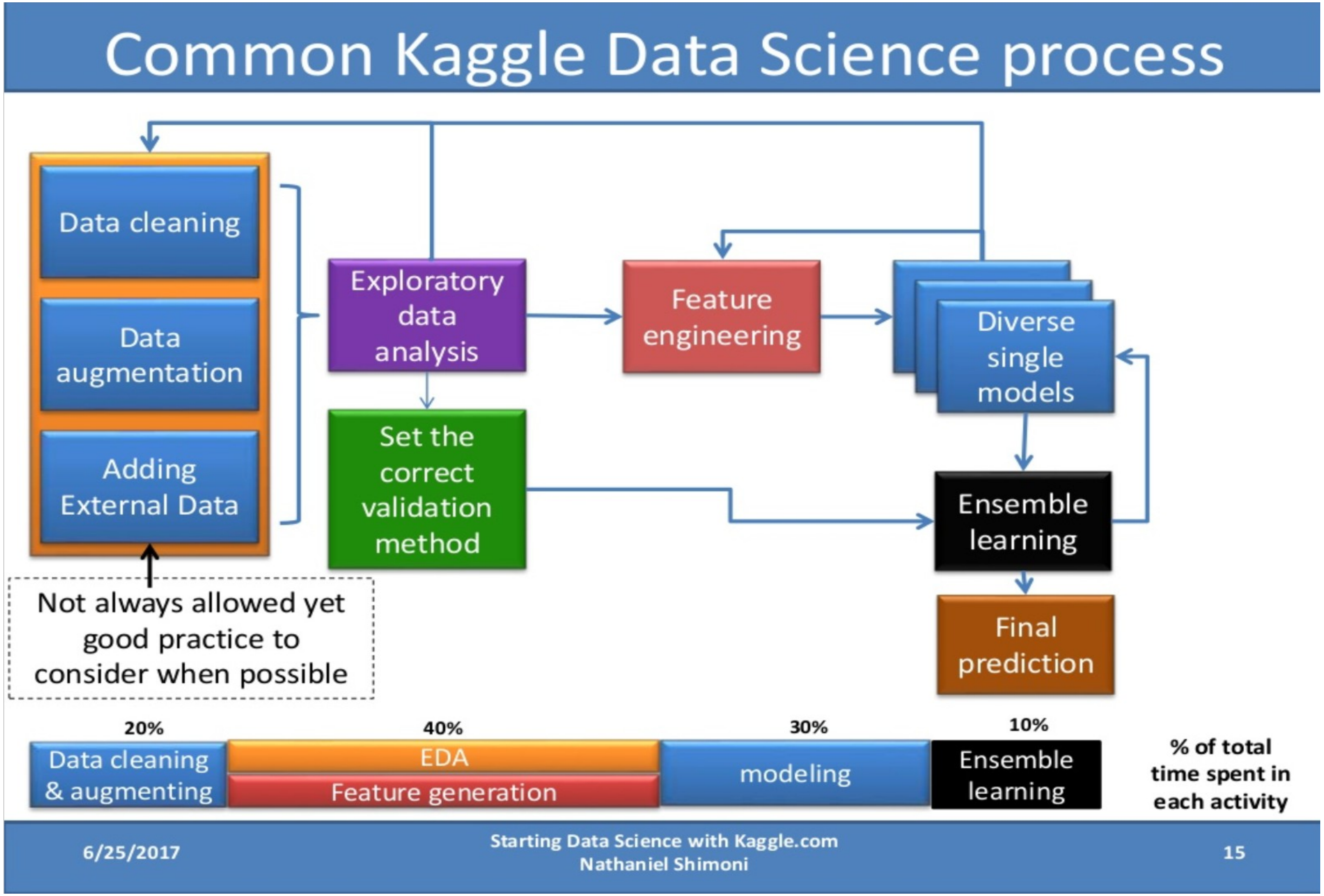
But, is human intelligence the ideal?

- The definition of intelligence is complicated
- In some aspects computers are already more “intelligent“
  - Computational speed, precision,
  - Memory, communication, reaction times,
  - Consistency, availability, etc.
- Depends on the application and goal
- Should computers imitate a human when solving a problem?..

Outline

- What is Machine Learning (ML)
  - Definition, history, ML and AI
  - ML and data science
- ML types and basic setup
  - Optimal decision boundary
  - Curse of dimensionality

What Kagglers say



https://www.kaggle.com/
ML competitions

https://www.slideshare.net/NathanielShimoni/starting-data-science-with-kagglecom

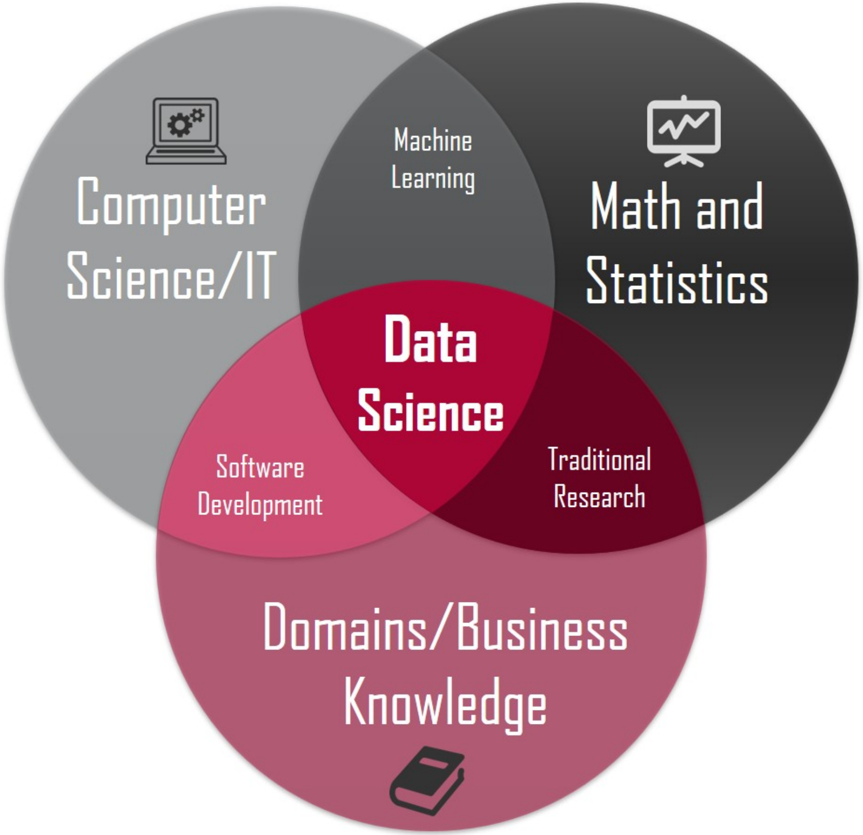
Machine learning in real world

- Like going to an expedition vs. racing in a track (Kaggle)
- A lot of work goes into extracting, preparing, understanding the data
  - Data engineering:
    - Data is usually messy: buried, errors, missing values, no standards…
    - Complicated by: big data, continuous streams, heterogeneous systems
  - Exploratory data analysis
  - Feature engineering...

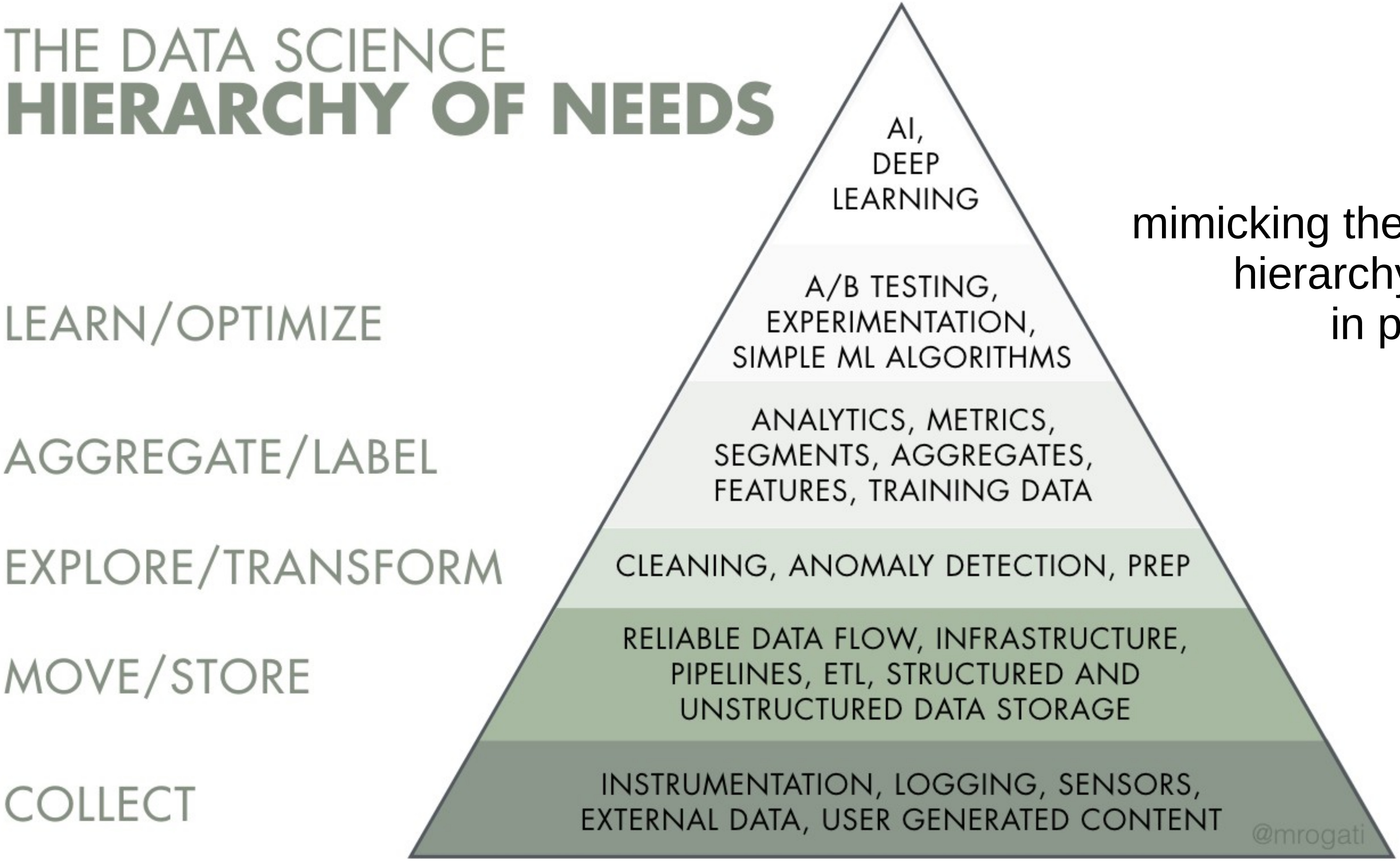
Machine Learning vs. Data Science

- Data Science, Data Analytics, Data Mining, etc. are more about humans understanding the data
  - Business intelligence/analytics even narrower
- ML is more about creating an automated system
- Help each other
  - Understand the data ↔ build/train ML models
- No clear-cut boundary

Data science: a popular definition



pic. from https://towardsdatascience.com/introduction-to-statistics-e9d72d818745



mimicking the Maslaw’s
hierarchy of needs
in psychology

https://hackernoon.com/the-ai-hierarchy-of-needs-18f111fcc007

Exploratory data analysis

- To understand the data, assess quality
- A lot of statistics, visualizations, looking at the data from different angles
- Good tools:
  - Jupyter, Pandas, Matplotlib/Seaborn, …
- Finding anomalies, fixing, cleaning the data
- May involve simple ML models, including unsupervised

Outline

- What is Machine Learning (ML)
  - Definition, history, ML and AI
  - ML and data science
- ML types and basic setup
  - Optimal decision boundary
  - Curse of dimensionality

Machine Learning: what

- **Supervised** – Learning by close example
  → We know the correct answer, Error minimization
- **Unsupervised** – “making sense” of data
  → A representation minimizing some property(-ies)
- **Reinforcement** – reward/punishment
  → Exploration and exploitation
- **Competition** – co-evolution
  → Evolutionary algorithms
  → Adversarial training
- **Semi-supervised, self-supervised, etc.** – mixed/other setups

Machine Learning: when

- **Offline** machine learning
  - Learning and exploitation phases are separate
  - The model is frozen after learning
- **Online/continuous/lifelong** machine learning
  - Learning always continues
  - ✔ Adapts to new conditions
  - ✗ May loose previous knowledge

Mathematical abstractions

- Variables
  - scalar a
  - vector $\mathbf{a}$
  - matrix, tensor $\mathbf{A}$
- Function $b=f(\mathbf{a})$
- Sets $\{a_i\},\ i=1,\ldots,k$

  $a\in\{0,1\}$ $\mathbf{a}\in\{0,1\}^n$

- Set of real numbers
  - $a\in\mathbb{R}$
  - $\mathbf{a}\in\mathbb{R}^n$
  - $\mathbf{A}\in\mathbb{R}^{n\times m}$
- A vector $\mathbf{a}\in\mathbb{R}^n$ is a point in an n-dimensional space

Supervised Machine Learning: Data

- (Training) Data $\{(\mathbf{x}, \mathbf{y})^{(i)}\},\ i=1,\ldots,m$
  or $(\mathbf{X}, \mathbf{Y}),\ \{(\mathbf{x}^{(i)}, \mathbf{y}^{(i)})\}$
  - Inputs $\mathbf{x}^{(i)} \in \mathbb{R}^{n_x},\ \mathbf{X} \in \mathbb{R}^{n_x \times m}$
  - Desired "target" or "teacher" outputs

    $\mathbf{y}^{(i)} \in \mathbb{R}^{n_y},\ \mathbf{Y} \in \mathbb{R}^{n_y \times m}$

We will typically talk about vectors as data points $(\mathbf{x}, \mathbf{y})$ for equation simplicity

$(\ \mathbf{x}\ ,\ \mathbf{y}\ )$

( $,(0,1))$

( $,(1,0))$

( $,(1,0))$

( $,(0,1))$

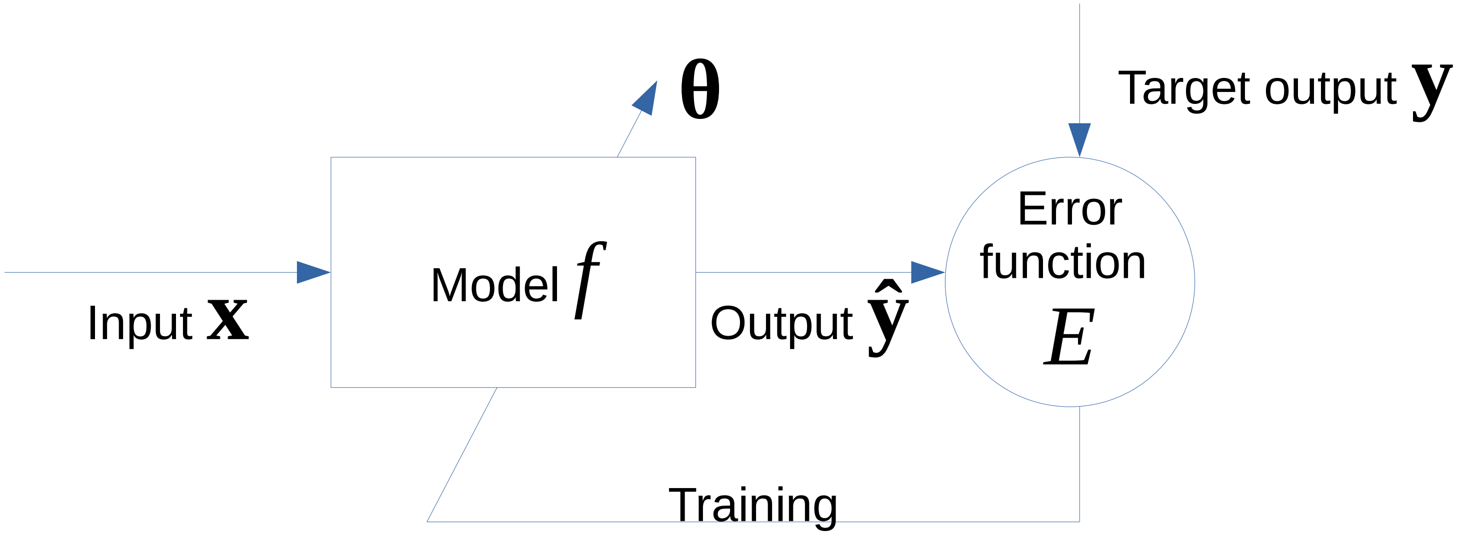
cat dog

Supervised Machine Learning Setup

- (Training) Data $\{(\mathbf{x},\mathbf{y})^{(i)}\},\ i=1,\ldots,m$ or $(\mathbf{X},\mathbf{Y}),\ \{(\mathbf{x}^{(i)},\mathbf{y}^{(i)})\}$
  - Inputs $\mathbf{x}^{(i)}\in\mathbb{R}^{n_x},\ \mathbf{X}\in\mathbb{R}^{n_x\times m}$
  - Desired “target” or “teacher” outputs $\mathbf{y}^{(i)}\in\mathbb{R}^{n_y},\ \mathbf{Y}\in\mathbb{R}^{n_y\times m}$
- Error (“loss”, “cost”) function $E(\mathbf{Y},\hat{\mathbf{Y}})$
  - Between all desired targets $\mathbf{Y}$ and model outputs $\hat{\mathbf{Y}}$
  - E.g., Mean Square Error $E_{MSE}(\mathbf{Y},\hat{\mathbf{Y}})=\frac{1}{m}\sum_{i=1}^{m}|\mathbf{y}^{(i)}-\hat{\mathbf{y}}^{(i)}|^2=\frac{1}{m}\sum_{i=1}^{m}\sum_{j=1}^{n_y}(y_j^{(i)}-\hat{y}_j^{(i)})^2$
- Model $\hat{\mathbf{y}}=f_{\boldsymbol{\theta}}(\mathbf{x})=f(\mathbf{x},\boldsymbol{\theta})$
  - Has parameters $\boldsymbol{\theta}$
- Training algorithm
  - Tunes the model parameters $\boldsymbol{\theta}$ to minimize E

Training as Optimization

- Training: $\boldsymbol{\theta} = \arg\min_{\boldsymbol{\theta}} E\left(\mathbf{Y}, f\left(\mathbf{X}, \boldsymbol{\theta}\right)\right)$



Good Model

- Generic/universal
  - Is specialized by the parameters
- High expressive power
  - Power (number of parameters) tunable by meta- or hyper-parameters
- Computationally efficient
- Easy to train w.r.t. relevant Es
- There are many in ML, with quite different properties

- (or domain-specific)

Supervised ML task types

- Static $\{(\mathbf{x},\mathbf{y})^{(i)}\},\ i=1,\ldots,m$
  model: $\hat{\mathbf{y}}=f(\mathbf{x},\boldsymbol{\theta})$
  - Classification
    $y\in\{0,\ldots,k-1\}$
    or $\mathbf{y}\in\{0,1\}^k=\{0,1\}^{n_y}$
  - Regression
    $\mathbf{y}\in\mathbb{R}^{n_y}$
  - ...
- Temporal $(\mathbf{x}(n),\mathbf{y}(n)),\ n=1,\ldots,T$
  $\hat{\mathbf{y}}(n)=f(\ldots,\mathbf{x}(n-1),\mathbf{x}(n),\boldsymbol{\theta})$
  - Detection $y(n)\in\{0,1\}$
  - Classification $\{(\mathbf{x}(n),\mathbf{y})^{(i)}\}$
  - Pattern recognition $\mathbf{y}(n)\in\{0,1\}^{n_y}$
    (= detection + classification)
  - Prediction $\mathbf{y}(n)=\mathbf{x}(n+1)$
  - Pattern generation, ...

Outline

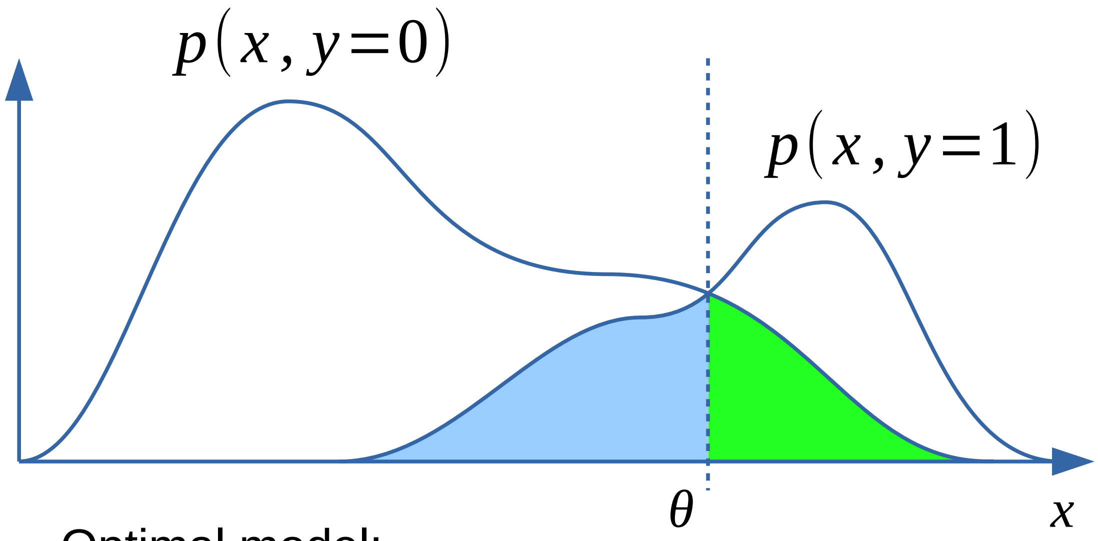
- What is Machine Learning (ML)
  - Definition, history, ML and AI
  - ML and data science
- ML types and basic setup
  - Optimal decision boundary
  - Curse of dimensionality

Optimal Decision Boundary

A very simple case:

$$\mathbf{x} = x \in \mathbb{R}, \quad \mathbf{y} = y \in \{0,1\}$$

- Estimate joint PDF
  - from histograms
  - or prior + parameters
- Put the decision boundary where PDF's intersect
  - To minimize misclassification



Optimal model:

$$\hat{y} = f(x, \theta) = H(x - \theta) = \begin{cases} 0, & \text{for } x < \theta \\ 1, & \text{for } x \geq \theta \end{cases}$$

(Heaviside step function)

Probabilistic View on ML

- Accurate estimation of the joint probability density (or distribution) $p(\mathbf{x},\mathbf{y})$ is all we would
  - need
    - All answers can be computed as marginal PDF's,
      - $p(\mathbf{y}|\mathbf{x})$ would more than make up for $\hat{\mathbf{y}}=f(\mathbf{x})$

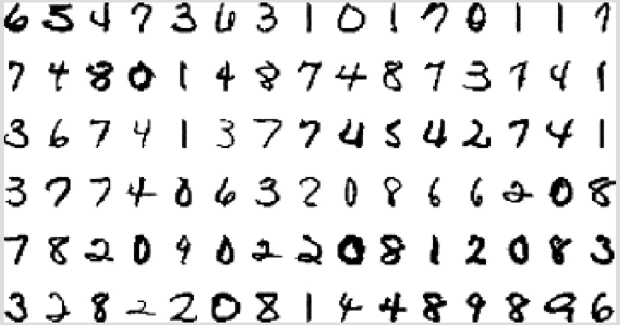
        a "degenerate" case of distribution
  - hope to get
    - There is no more knowledge about the task in the data

Outline

- What is Machine Learning (ML)
  - Definition, history, ML and AI
  - ML and data science
- ML types and basic setup
  - Optimal decision boundary
  - Curse of dimensionality

Classical Task: Digit recognition

- USPS



Pic from: Z. Zhenga, J Yang, 06

- MNIST

https://en.wikipedia.org/wiki/MNIST_database

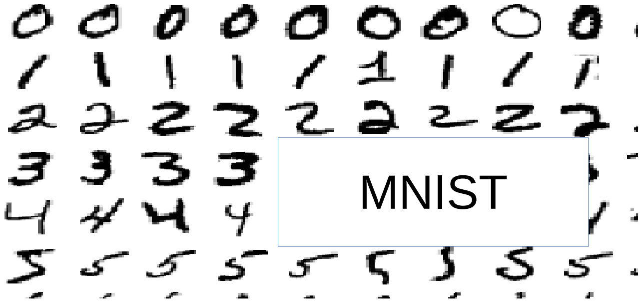
- Google house numbers

http://ufldl.stanford.edu/housenumbers/

Same Approach?

- $28 \times 28 = 784$ gray-scale pixels

  $\mathbf{x} \in \mathbb{R}^{784}, \quad \mathbf{y} = y \in \{0, 1, \ldots, 9\}$



Pic from: H. Hoffmann, 05

- Compute histogram?
  - l bins per dimension $\rightarrow l^{784}$ bins
  - $l = 2 \rightarrow 2^{784} > 10^{236}$ bins!
  - Almost all empty...

… it is estimated that the there are between 10^{78} to 10^{82} atoms in the known, observable universe.

https://www.universetoday.com/36302/atoms-in-the-universe/

– Fail!

Curse of Dimensionality

- For high dimensional data it's impossible to estimate probability densities
- *A few lonely samples are scattered in a huge wasteland of space...*

- Most real world data are high-dimensional and suffer from this

https://en.wikipedia.org/wiki/Curse_of_dimensionality

Questions?

End of part 2 /8

▮▮▯▯▯▯▯▯

To be continued…

mantas.lukosevicius@ktu.lt